

WHAT IS CLAIMED IS:

1. A method of aligning a query protein sequence with a template consisting of a set of pre-selected protein structures in a database, comprising the steps of:
selecting an energy function, said energy function being a sum of energy parameters and weighting factors;
determining values for weighting factors in said energy function;
establishing linear programming (LP) constraints for threading (or aligning) said query protein sequence with each structure in said set of pre-selected protein structures in a database; and
performing a linear programming analysis based on a linear programming formulation including said energy function under said constraints, to optimally align said query protein with said template.
2. The method of claim 1, wherein said step of determining weighting factors comprises the step of using training to determine values for said weighting factors.
3. The method of claim 2, where said energy function comprises the function: $\min W_m E_m + W_s E_s + W_p E_p + W_g E_g + W_{ss} E_{ss}$.
4. The method of claim 1, where alignment gaps are confined to loops.
5. The method of claim 1, where only interaction between core residues is considered.
6. The method of claim 1 wherein said step of performing a linear programming analysis is done on the assumption that solutions are likely to be integral.
7. The method of claim 6, wherein said step of performing a linear programming analysis comprises the step of using a branch and bound technique to perform said linear programming analysis.

8. The method of claim 1, where said linear programming constraints comprise Constraints (8) - (15).
9. The method of claim 1, where said linear programming constraints comprise Constraints (8), (14), (15) and (16) - (19).
10. The method of claim 1, where said linear programming constraints comprise Constraints (8), (14), (15), (20) and (21).
11. The method of claim 1 further comprising the step of performing graph reduction to decrease the number of integer variables and speed up the LP analysis.
12. The method of claim 1, further comprising the step of performing fold analysis using a support vector machine (SVM) algorithm.
13. The method of claim 1, comprising step of generating a dense contact graph prior to said step of performing a linear programming analysis.
14. A method of alignment comprising the steps of:
formulating the protein threading problem as a large scale integer programming problem;
relaxing this problem to a linear programming problem; and solving the integer program by a branch-and-bound method.
15. A system for aligning proteins comprising:
a computer operable to align a query protein sequence with a template consisting of a set of pre-selected protein structures in a database, by performing the steps of:
selecting an energy function;
determining values for weighting factors in said energy function;
establishing linear programming (LP) constraints for threading (or aligning) said query protein sequence with each structure in said set of pre-selected protein structures in a database; and

- 25 -

performing a linear programming analysis based on a linear programming formulation including said energy function under said constraints, to optimally align said query protein with said template.